

Article

Chemical Source Localization Fusing Concentration Information in the Presence of Chemical Background Noise [†]

Víctor Pomareda ^{1,2}, Rudys Magrans ¹, Juan M. Jiménez-Soto ¹, Dani Martínez ³, Marcel Tresánchez ³, Javier Burgués ^{1,2}, Jordi Palacín ³ and Santiago Marco ^{1,2,*}

¹ Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia, Baldiri Reixac 4-8, Barcelona 08028, Spain; vpomareda@ibebarcelona.eu (V.P.); rmagrans@ibebarcelona.eu (R.M.); jmjimenez@ibebarcelona.eu (J.M.J.-S.); jburgues@ibebarcelona.eu (J.B.)

² Department of Engineering: Electronics, Universitat de Barcelona, Martí i Franqués 1, Barcelona 08028, Spain

³ Department of Computer Science and Industrial Engineering, Universitat de Lleida, Jaume II 69, Lleida 25001, Spain; dmartinez@diei.udl.cat (D.M.); mtresanchez@diei.udl.cat (M.T.); palacin@diei.udl.cat (J.P.)

* Correspondence: smarco@ibebarcelona.eu; Tel.: +34-934-029-070

[†] This paper is an extended version of our paper: Pomareda, V.; Marco, S. Chemical Plume Source Localization with Multiple Mobile Sensors using Bayesian Inference under Background Signals. In Proceedings of the International Symposium on Olfaction and Electronic Nose, New York, NY, USA, 2–5 May 2011.

Academic Editor: W. Rudolf Seitz

Received: 14 February 2017; Accepted: 11 April 2017; Published: 20 April 2017

Abstract: We present the estimation of a likelihood map for the location of the source of a chemical plume dispersed under atmospheric turbulence under uniform wind conditions. The main contribution of this work is to extend previous proposals based on Bayesian inference with binary detections to the use of concentration information while at the same time being robust against the presence of background chemical noise. For that, the algorithm builds a background model with robust statistics measurements to assess the posterior probability that a given chemical concentration reading comes from the background or from a source emitting at a distance with a specific release rate. In addition, our algorithm allows multiple mobile gas sensors to be used. Ten realistic simulations and ten real data experiments are used for evaluation purposes. For the simulations, we have supposed that sensors are mounted on cars which do not have among its main tasks navigating toward the source. To collect the real dataset, a special arena with induced wind is built, and an autonomous vehicle equipped with several sensors, including a photo ionization detector (PID) for sensing chemical concentration, is used. Simulation results show that our algorithm, provides a better estimation of the source location even for a low background level that benefits the performance of binary version. The improvement is clear for the synthetic data while for real data the estimation is only slightly better, probably because our exploration arena is not able to provide uniform wind conditions. Finally, an estimation of the computational cost of the algorithmic proposal is presented.

Keywords: machine olfaction; odor robots; chemical sensors; Bayesian inference

1. Introduction

Localization of chemical sources in urban scenarios (large cities) is a major challenge for intelligence and police authorities. In the clandestine production phase of illicit substances like explosives, but also drugs, significant levels of precursors are spread in the atmosphere. Such suspicious compounds could be reported by a system of mobile sensors and could be located using

localization algorithms, providing complementary information to the authorities for intervening at an early stage.

Several strategies for source localization have been proposed in the literature [1]. These strategies have been integrated into robot systems with tracking abilities. Many of these tracking strategies have been inspired by bacteria or animal behavior using olfaction for foraging or mating: lobster, blue crabs, ants and moths provide behavioral models for odor tracking [2,3]. One of the simplest strategies consists in seeking changes in local concentration within an odor plume assuming a smooth chemical gradient in a diffusion dominated flow. However, this approach, called chemotaxis [4], is not useful in a realistic environment where fluid flow is dominated by turbulence, which can be caused by forced ventilation, temperature gradients or the presence of obstacles. Under these conditions, on the order of ten min are required to determine the time-averaged concentration with sufficient accuracy to perceive the gradient concentration [5]. In consequence, the exploration of the area of interest becomes too slow. Some other strategies additionally exploit both fluid velocity information and chemical concentration (anemotaxis [6,7]). More recent proposals based on information theory, like infotaxis [8] are based in binary detections and, information plays a role similar to that of concentration in chemotaxis. Odor patches are expected to be found only intermittently in the medium, and then information is sparse. In some cases, methods aim to estimate the gas distribution through analytical Gaussian models [9–11], others are focused on to create a plume mapping via hidden Markov methods [12], whereas in Farrell et al. [13] a strategy for chemical plume tracing and source location declaration is presented.

Navigation experiments aiming to find chemical sources are strongly limited by the limit of detection and selectivity of the low cost chemical sensors and even medium-priced detectors (e.g., ion mobility spectrometers). Thus, the rapid decay in the chemical concentration with increasing distance from the source can be a critical issue. Poor limits of detection result in a reduced area where the plume can effectively be detected. This is especially important in applications where the search zone has an area of several square kilometers. In such situations, it becomes very important to set the detection thresholds very close to the noise level, but this would result in a high number of false alarms and most localization algorithms would fail catastrophically. To the best of our knowledge, none of the published methods have addressed this problem.

Additionally, in any real scenario, there could be background levels of a multitude of chemicals caused by environmental pollution. Because of limited selectivity, there will be substances which will produce interference in the detector reading, hindering the detection and localization tasks. The presence of combination of detector electronic noise and mainly interfering chemical agents result in variable background readings that may change with time and with the position of the detector. These shifts in background levels hamper the selection of an optimum threshold that is usually considered to be constant all along the area under exploration. As far as we know, this problem has not been previously tackled in the literature.

To address these two issues (threshold close to the limit of detection and presence of background levels) probabilistic approaches like plume mapping Bayesian methods appear to be a good choice. Pang and Farrell published a source-likelihood mapping approach based on Bayesian inference in 2006 [14]. The main idea behind the algorithm consists in implementing a stochastic approach for plume modeling and in estimating the most likely source position considering the sequence of detection/non-detection events and fluid flow measurements along the robot's trajectory. Pang's algorithm has been tested successfully with data previously collected using an autonomous underwater vehicle [13].

However, this algorithm uses binary detection events, and no chemical concentration information is used to build the probability map, since it only considers the concentrations above a certain threshold as detection or non-detection events. Moreover, after setting the threshold level, the approach assumes that the rate of false alarms is very low. In a real scenario where background signals are present, this is only achieved when the threshold is set at a high level. However, this option seriously reduces the

maximum plume detection distance. Therefore, there is a trade-off; on the one hand, the threshold needs to be set low enough (close to the sensor detection limit) if chemicals from the source are to be detected at large distances; on the other hand, the threshold needs to be high enough to prevent false alarms. So, how to set the threshold level becomes a critical issue in real environments using existing approaches, especially when the background intensity is non-uniform in the explored area.

Finally, in order to speed up the area exploration it is important that the algorithm can be extended to work with multiple robots. Recently, Kang and Li [15] have presented a novel plume tracking algorithm via multiple autonomous robots by using a leader-follower strategy, demonstrating its superiority versus a single robot algorithm, in terms of both the computational cost and the accuracy in source location. Meng et al. [3] have also studied the multi-robot problem for plume tracking in fluctuating airflow environments, showing the efficiency and robustness of the adapted ant colony optimization (ACO) algorithm over the traditional ACO algorithm. Meng remarks the importance of a proper number of robots and a well-defined cooperation mechanism, although it is not strictly necessary to track the plume to obtain a good estimation of the source location. The most likely source position is estimated during the robot's mission for arbitrary trajectories by recursively building a probability map using Bayesian inference. The problem of background estimation over the exploration area can be considered as the problem of scalar field mapping. Algorithmic approaches using mobile sensor networks have been already proposed, however they assume that the agents have communication capabilities so that their exploration paths are optimized after data fusion [16–18]. However, in our scenario we expect that the sensors can be mounted on vehicles that serve other tasks beyond chemical source localization. In this sense, we consider cases where there is no feedback between the chemical sensing and the agent's trajectories.

Our main motivation for the present work is to extend the Bayesian plume source localization algorithm, previously described by Pang and Farrell [14], using the chemical concentration (instead of binary detections) and assessing its performance in simulated and real environments, where background signals may arise. Thus, Pang's algorithm is reformulated for use with continuous analog concentration readings instead of binary detections. Moreover, the algorithm is extended to work with multiple mobile sensors. This new approach requires a probabilistic model for the background and for the plume which are described in the following sections.

The present algorithm was initially developed [19] for applications to Home Security (European project LOTUS: Localization of Threat Substances in Urban Society FP7-SEC-217925). In the considered scenario, police cars equipped with GPS and gas chemical sensors (eventually other fast analyzers like ion mobility spectrometers can be used) carry out their routine patrols while sending the sensors' readings to a central station. Instead of moving towards the chemical source, the patrol vehicles would maintain their normal patrol routes while a centralized system is continuously analyzing the acquired signals seeking suspicious activity.

The paper is organized as follows: Sections 2.1 and 2.2 and shows the basics of the algorithm and the plume and background model. Sections 2.3 and 2.4 describes the synthetic and real scenarios for the test and Section 3 reports the results and the analysis.

2. Materials and Methods

2.1. Stochastic Models for Plume and Background

Our proposal requires making some assumptions about the dispersion of the plume (the stochastic model in Section 2.1.1) and a background model (Section 2.1.2). Given an instantaneously measured concentration c , it is assumed that there are two additive contributions: one due to the background (c_b) and one due to the plume (c_p), thus:

$$c = c_b + c_p \quad (1)$$

Since we consider all these concentrations as random variables, the probability density function (PDF) of the measured concentration will be the convolution of the PDFs of the two additive terms. In the next sections, we will describe how we model the probability density of both terms.

2.1.1. Stochastic Model for Chemical Concentration Measurements

The basis of our stochastic model [20] for the chemical plume is the analytical Gaussian plume model (GPM) [21,22]. Time-averaged plume concentration follows a Gaussian distribution lengthways to the flow direction if the time average is at least 10 min, as demonstrated in previous works [23–25]. This model has been widely used for its simplicity and is appropriate when dispersion is governed by atmospheric turbulence under uniform wind conditions. Atmospheric turbulence is determined by the stability of the atmosphere and the height above the surface layer [22]. The basic expression for the GPM under a continuous release is:

$$\bar{c}(x, y, z) = \frac{q}{2\pi U_a \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{(h-z)^2}{2\sigma_z^2}\right) \quad (2)$$

where \bar{c} is the time averaged concentration (in g/m^3) in a location with coordinates x (downwind), y (crosswind) and z (vertical); q is the continuous source release rate or source strength (in g/s); U_a is the mean wind speed in the downwind direction (in m/s); h is the plume height (in m); and σ_y, σ_z are the dispersion coefficients (in m), in crosswind and vertical direction respectively, modeled as: $\sigma_y = a \cdot x^b$, $\sigma_z = c \cdot x^d$, where a, b, c and d are parameters obtained from a table [22] and their values depend on the atmospheric conditions which can be organized in six levels (from A-very unstable to F-very stable) according to the Pasquill's stability classes [26]. In Equation (2), the resulting concentration distribution is due to the transport of chemicals by advection (due to the mean wind speed), due to concentration gradients within the plume width (lateral dispersion due to diffusion, but also turbulent mixing) and due to plume meandering. The decay of mean concentration is exponential, thus concentration levels below the sensor detection limit are very quickly achieved. This issue makes the setting of the threshold level critical; especially, if the source should be detected far from the release point.

The GPM considers the time-averaged characteristics of a plume dispersed in a turbulent flow, but the sensors will be responding to the instantaneous plume characteristics (here we assume that the chemical sensor dynamics are much faster than 10 min, which is usually the case). For short time-scale studies, the chemical puff movement can be modeled as a random walk (because of transversal velocity fluctuation) overlapped on the downflow advection (because of mean velocity) [27]. However, we propose an alternative approach, for which an additional component needs to be added to the GPM to model the unpredictable and random fluctuations in concentration due to turbulent stirring and plume meandering. Yee et al. [28–30] have carried out empirical studies on plume statistics in urban areas using scale fluid models in a variety of plume conditions and urban geometries. Their results prove that instantaneous concentration fluctuations fit very well the clipped-gamma PDF over a very wide range of atmospheric conditions and at several receptor positions [30]. The clipped-gamma distribution (CGD) is defined in terms of four parameters γ, k^*, s and λ as [28]:

$$f(\hat{c}) = \left(\frac{\hat{c} + \lambda}{s}\right)^{k^*-1} \frac{\exp(-(\hat{c} + \lambda)/s)}{s\Gamma(k^*)} + (1 - \gamma)\delta(\hat{c}) \quad (3)$$

where \hat{c} is the instantaneous concentration, $\Gamma(\hat{c})$ is the gamma function, $\delta(\hat{c})$ is the Dirac delta function and λ, k^* and s are the shift, shape and scale parameters, respectively. The total PDF is composed of a mixed fluid part due to in-plume mixing of eddies containing the target substance (the first term on right-hand side), and an unmixed ambient fluid part (the second term on right-hand side) caused by plume meandering which produces intermittent periods of zero concentration for a fraction of time $(1 - \gamma)$, being γ the intermittency factor. Although Equation (3) is specified in terms of four parameters, it can be uniquely modeled by the mean (M) and the standard deviation (SD) of a series

of readings [28]. There is a simple relation among M , SD , and the plume intermittency; thus, γ is determined as:

$$\gamma = \gamma(k^*, s, \lambda) = \frac{\Gamma(k^*; \lambda/s)}{\Gamma(k^*)} = \min\left(1, \frac{3}{K_{int}^2 + 1}\right) \quad (4)$$

where $\Gamma(v; c)$ corresponds to the incomplete gamma function and K_{int} is the ratio between the SD and the mean of the series of readings at a fixed position. For a specified value of K_{int} , the parameters λ , k^* and s can be obtained solving a set of transcendental equations, thus making Equation (3) totally defined. The details for computing these parameters are given in [28].

While the mean concentration decreases rapidly in the downwind direction (see Equation (2)), the magnitude of the fluctuations decreases even more rapidly. As described by Webster [5], K_{int} is estimated to decrease as x^θ (where x is the distance from the source and $\theta < 0$), being roughly K_0 times the mean time-averaged concentration at a certain distance x_0 from the source. Therefore, this parameter can be modeled as:

$$K_{int} = K_0 \cdot \left(\frac{x}{x_0}\right)^\theta \quad (5)$$

To model the instantaneous concentrations due to a chemical plume at a certain distance from the source, the clipped-gamma distribution is used (Equation (3)). The mean value M of the series of concentrations due to the plume is related to the GPM (Equation (2)) and the SD can be modeled as:

$$\sigma_{SD} = K_0 \cdot \left(\frac{x}{x_0}\right)^\theta M \quad (6)$$

Since the PDF depends on the mean and the SD , and these parameters depend on the distance from the source, the PDF of the instantaneous readings contains information about the relative position between the sensor and the source. Since concentrations fluctuations (intermittencies) decrease faster than the mean value with the downwind distance to the source, the plume becomes homogeneous faster than the mean concentration dilutes [28,31].

Additionally, previous literature has characterized the power spectral density (PSD) of concentration readings within a dispersing plume. We will follow the model described by Jones et al. [31].

2.1.2. Stochastic Background Model

In real scenarios such as residential areas or urban environments, pollution or interfering substances are expected to be found. This problem becomes even more serious due to the common use of partially selective sensors such as Metal Oxide Sensors (MOXs), or Photoionization detectors (PIDs).

Moreover, meteorological conditions (wind conditions and atmospheric stability) could change within a timescale of several hours, or there might be changes in polluting emissions due to day-night cycles of human activity (including motor vehicles or factories) [32–34]. Because of these reasons the background can be considered to change slowly.

As it has been previously described [5], the dispersion of a chemical plume in a turbulent flow shows a highly intermittent nature with background or zero concentration for long periods of time separated by high peaks of concentration. This behavior will help us to estimate the background model using robust measures that reject the plume peaks.

For the implementation of the proposed algorithm the exploration area is divided into a uniform grid of rectangular cells. The algorithm estimates the background probability distribution at each cell of the grid. The background is modelled as a Gamma PDF where the standard deviation is smaller than the mean. In those conditions the Gamma function resembles a Gaussian PDF but defined only for positive values of the random variable concentration. We will consider that the PSD of the background is bandlimited white noise.

2.2. Estimation of the Models From the Concentration Readings

2.2.1. Bayesian Estimation of the Likelihood Map for Chemical Source Presence

A summary of the notation used for the description of the algorithm can be found in Table A1. The search area contains N_c rectangular cells of size $L_x \cdot L_y$ each, where L_x and L_y are the cell lengths in the x and y directions of the grid map, respectively. The size of these cells is a trade-off between the spatial resolution of the algorithm and the computational cost.

Let $0 \leq \alpha'_i \leq 1$ represent the probability that the chemical source is in cell i . It is assumed that the search area contains exactly one source, hence $\sum_{i=1}^{N_c} \alpha'_i = 1$. It is assumed that the prior information on the potential existence of a chemical source is given by previous intelligence research. We do not consider in this formalism the case where there is uncertainty on the presence or not of a chemical source.

Initially (at $t = t_0$), if no information about the source location is available, all cells are initialized to be equally likely to contain the chemical source: $\alpha'_i(t_0) = \left(\frac{1}{N_c}\right); \forall i \in [1, N_c]$.

Given that we measure a concentration at time t_k in cell j , i.e., c_j refers to a concentration reading in cell j , we can calculate the source probability map based on this single reading. A Bayesian approach is used to determine if the main contributor to the measurement is the background or the presence of a plume patch. The posterior probability for the presence of a plume, given the measurement, is calculated using Bayes' theorem:

$$P(A|c_j) = \frac{P(c_j|A) \cdot P(A)}{P(c_j)} \quad (7)$$

where A corresponds to the event "the concentration reading was caused by an emitting source upstream (plus a background level)". To infer where the source is located, the posterior probability of a source emitting (left-hand term in Equation (7)) can be further decomposed into the probability of that source being located at each cell. Again, using the Bayes' Theorem:

$$P(A_i|c_j) = \frac{P(c_j|A_i) \cdot P(A_i)}{P(c_j|A_i) \cdot P(A_i) + P(c_j|\bar{A}_i) \cdot [1 - P(A_i)]} \quad (8)$$

where \bar{A} means "the concentration reading was caused only by background levels and not by an emitting source", subindex i refers to a source located in cell i and subindex j refers to the current cell j where the measurement was taken.

Since we consider that a background of interfering substances is always present, in the absence of plume only the background component (c_b) is present; in the presence of plume, both components (c_b and c_p) are present and the concentrations are modeled by the convolution of the plume PDF and the background PDF. However, even in the presence of the plume c_p may be zero due to plume intermittency. The PDF of the concentration component due to the source is modeled using the GPM for the means, and SD from Equation (6) considering the relative position of cell i (potential source location) and cell j (location of the sensor).

Taking these considerations into account, the previous probabilities have the following interpretation: $P(A_i)$ is the prior probability of the presence of a source at the cell i , $P(c_j|A_i)$ is the probability that the measurement at cell j is due to addition of the background at cell j and a plume due to a source at cell i , and it is obtained by the convolution between the PDF of the plume and the PDF of the background at cell j . $P(c_j|\bar{A}_i)$ is the probability that the measurement of chemicals at cell j is due to the current background at cell j , and it is obtained from the PDF of the background at cell j . We define $S_{ij}(t_k) \equiv P(A_i|c_j)$ as the probability of having a source in cell i given that a certain amount of chemical was measured at cell j at time t_k .

However, in this approach where we consider each cell independently, $\sum_{i=1}^{N_c} S_{ij}(t_k) = 1$ is not guaranteed. Therefore, the result is normalized to ensure the total probability is 1 when individual cell probabilities are added up. The nomenclature is henceforth the same as that used in [14]:

$$\beta_{ij}(t_k) = \frac{S_{ij}(t_k)}{\sum_{i=1}^{N_c} [S_{ij}(t_k)]} \quad (9)$$

Now $\sum_{i=1}^{N_c} \beta_{ij}(t_k) = 1$ is guaranteed and $\beta_{ij}(t_k)$ calculated over all cells ($i = 1, \dots, N_c$) gives the source probability map at time t_k based on a single measured concentration at cell.

Using Bayesian theory [20] and following the same procedure described in [14], each new measurement can be incorporated recursively to update the source probability map.

$\alpha'_{ij}(t_k) = P(A_i|B(t_k))$ is defined as the probability of cell i containing the source, given the sequence of concentrations $B(t_k)$ along the trajectory of the mobile sensors up to time t_k . Defining $P(A_i|D_j(t_k)) = \beta_{ij}(t_k)$, where $D_j(t_k)$ is the measured concentration at time t_k ; $P(A_i|B(t_k))$ is computed from $B(t_{k-1})$ and $D_j(t_k)$, which are supposed to be independent events, finally obtaining:

$$\alpha'_{ij}(t_k) = N_c \cdot \alpha'_{ij}(t_{k-1}) \cdot \beta_{ij}(t_k) \quad (10)$$

where, if $\alpha'_{ij}(t_k)$ is computed over all cells ($i = 1, \dots, N_c$), an updated source probability map is obtained recursively.

Independently of the number of mobile sensors, the only information required by the algorithm is: the position where the measurement was obtained, the chemical concentration reading and the fluid flow measurement. The extension to multiple mobile sensors is as follows. We just build an integrated sequence of readings by addressing in a circular manner the set of mobile sensors. Then we send this sequence of measurements to the original algorithm. In other words, we do not fuse a posteriori maps built from individual sensors. We fuse the sequence of measurements at the input of the estimation algorithm.

2.2.2. Background Estimation

Background stochastic model is built from the robust estimation of the means and the dispersion to reject the effect of the intermittencies. We use the median and the median absolute deviation (MAD) to build the background probability density model [35]. These parameters are estimated over a buffer of the last 50 measurements. These measurements can be spread out in several cells depending on the speed of the robot (typically less than 10). This is not a problem if the spatial variation of the background is sufficiently smooth.

To allow for the model to adapt to a slowly changing background the parameters of the model (mean and standard deviation) are filtered with an exponential moving average. This filter weights the current estimation of the background with the old one depending on the time distance. In this way, the system is able to forget the old values and adapt to new ones.

2.3. Description of the Synthetic Test Cases

2.3.1. Synthetic Scenario Description and Simulation

For realistic simulations, the scenario envisioned considers atmospheric plumes in an urban area encompassing hundreds of thousands of square meters containing only one chemical source. The sensors located in vehicles can transmit their current position in the grid (e.g., using a GPS sensor) together with the chemical sensor readings, and they will explore the search area by moving across the cells performing random exploration. It is assumed that each vehicle mounts a single sensor. It is considered that the main task of the vehicles is not that of tracking the plume; but

patrolling a certain area and simultaneously updating a probability map for the source location using the available information.

The exploration arena in the synthetic case is as follows. In order to test both algorithms, a synthetic scenario (grid size of 1 km × 1 km) is generated. The area is divided into cells of size 100 m × 100 m. A square sub-grid of lanes (100 m separation between lanes) is interlaced over the main grid (Figure 1). The sensors will randomly explore the area over this last sub-grid of lanes which simulate streets within an urban environment (Manhattan style). Using this configuration, the movement of the sensors is constrained to certain lanes over the main grid. A clandestine laboratory for home-made explosive production emits explosive precursors. In this scenario, we consider that a chemical source emitting with a source strength $q = 2.90$ g/s, is placed in the grid with the source at the position (440 m, 440 m), which corresponds to coordinates (5, 5) on the rectangular grid. We consider that the main dispersed substance is acetone (molecular weight: 58.08 g/mol) at one atmosphere pressure and 25 °C. This Gaussian plume distribution has been generated from Equation (2) with the plume being dispersed in a 2D plane at the same height as the sensors ($z = h = 2$ m). It is assumed that there is no deposition of the substance on surfaces. In the simulations, the wind field is constant with the wind speed at $U_a = 2.5$ m/s and the wind direction at 45°. The dispersion coefficients σ_y and σ_z depend on wind conditions and atmospheric stability which has been set to neutral ('D' on the Pasquill-Gifford scale [26]). Moreover, a mean background distribution is deployed over the area with a different mean level in each cell and with SD equal to 60% of the mean value in all cells (based on our own recorded data using a PID sensor measuring in Barcelona outdoors over a period of several hours).

Five mobile sensors with a constant velocity of 15 km/h sense the area continuously. The sampling period of the sensors is set to 3 s, the detection limit to 0.1 ppm and the sensor resolution to 0.01 ppm, which are realistic values for PID technologies. We assume that the response time of the sensor is much faster than the typical 10 min time-average considered in the Gaussian Plume Model (Equation (2)). The total simulation time was set to 300 min.

The mean background level is different in each cell, but is stationary over time. Series of concentration fluctuations are generated in each cell considering the wind field created, the atmospheric conditions and the background. The stochastic model of the plume concentrations has been already described in Section 2.1.1. However, here we give some additional technical details concerning the practical implementation.

Since we have defined the PDF and the PSD which will be used to model concentration fluctuations, we use the percentile transformation method (PTM) described by Papoulis [36] to generate a series of concentration fluctuations with the desired PDF and PSD.

Specifically, the procedure to generate realistic plume readings consists of the following steps: (i) generate a time series of Gaussian white noise, (ii) filter the previous time series with the designed FIR filter to achieve the desired PSD, (iii) and apply the PTM. This method is based on the following expression:

$$c_i = F_c^{-1}(F_z(z_i)) \quad (11)$$

where z_i is a random sequence of Gaussian white noise having the desired PSD with cumulative distribution function (CDF) $F_z(z)$; c_i is the sequence of realistic readings in the cell with CDF $F_c(c)$. This CDF corresponds to the clipped-Gamma CDF (Equation (3)):

$$F(c) \equiv \Pr(C \leq c) = \int_{0^-}^c f(c') dc' = 1 - \frac{\Gamma(k^*; (c + \lambda)/s)}{\Gamma(k^*)} \quad (12)$$

and F_c^{-1} is the inverse clipped-Gamma CDF. The clipped-Gamma PDF (and its CDF) depends on the distance to the source. Its two parameters, the mean and the SD, are obtained from Equation (2) and Equation (6), respectively. Subsequently, these parameters are used to compute γ , k^* , s and λ , as explained in detail in [28].

Background concentrations are simulated as white noise with Gamma PDF. The background concentrations in each cell are added to the time-series of plume readings to obtain the final concentration readings at each cell.

An example of the concentration signals delivered by the sensors in this simulation scenario is shown in Figure 2. It can be observed that the SD of the fluctuations decreases faster than the mean concentration, making it difficult to differentiate between plume and background far from the source.

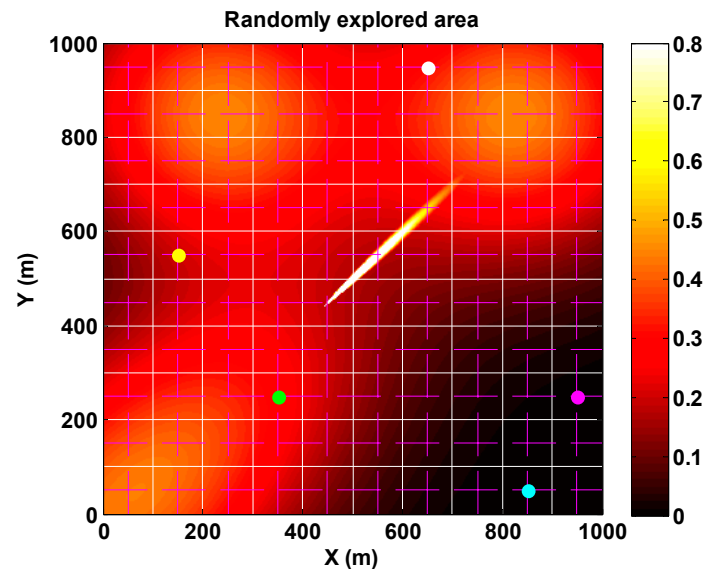


Figure 1. Randomly explored area ($1 \text{ km} \times 1 \text{ km}$) using five sensors (colored dots). The continuous white lines (-) define the cell boundaries. The sensors move in horizontal or vertical direction from the center of one cell to the center of a neighbor cell. The source is located at (440, 440) meters at cell (5, 5). Mean wind speed is set to 2.5 m/s and wind direction to 45° . Colored background refers to probability map at a particular time step.

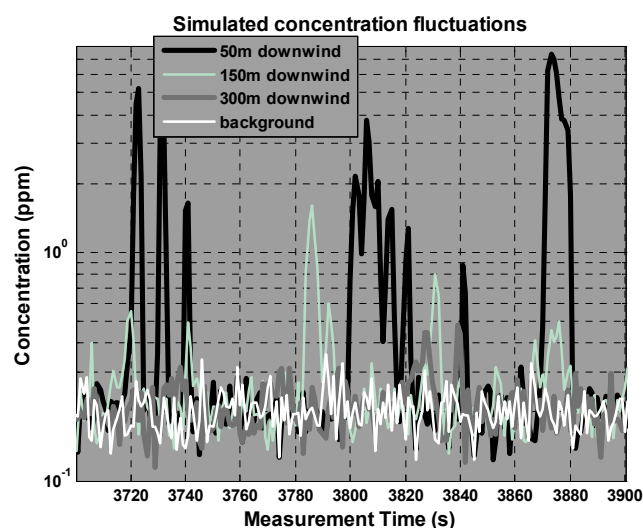


Figure 2. Simulated concentration fluctuations (under the specified conditions) at different fixed positions downwind from the source over a certain background level.

In the binary-based approach, it is assumed that the ratio of false alarms is very low, but the ratio of missed detection can potentially be very high, thus we define $\mu = 0.3$ (70% missed detections). However, the value of this parameter was not defined in the original work [14].

In the concentration-based approach, the source strength (q) should be initially guessed (this is typically done using previous information about the type of chemical source under investigation: clandestine lab, industrial toxic emissions, pollution, etc.). However, since it is almost impossible to guess this parameter with accuracy, we have studied the sensitivity of the algorithm to errors in this parameter. We scanned the guessed source strength two orders of magnitude around the central exact value. For the simulation studies presented in this work, the parameters used in Equation (6) are: $K_0 = 2.5$, $x_0 = 50$ m and $\theta = -0.75$, taking into account previous studies [5].

2.3.2. Synthetic Test Case 1: Behavior of the Binary Detector Algorithm Depending on the Background Level and Detector Threshold

The first simulation aimed to characterize binary-based approach when changing the concentration threshold that fires the detector signal. In order to investigate the robustness of the method against chemical noise, two spatially uniform background levels were considered (mean values: 0.05 ppm and 0.45 ppm). In both cases the standard deviation was 1/5 of the mean value.

A background level is low or high depending on the source strength; therefore, studying the case where the background is low is equivalent to saying that the source is potent, and studying the case where the background is high is equivalent to saying that the source is weak.

Ten random sensor trajectories were simulated for each threshold level in the binary detector (30 values in the linear range between 0.05 ppm and 3 ppm). For each trajectory, the probability at the real source location is assigned by the algorithm after 300 min of exploration time. We took as a figure of merit the mean probability averaging over all trajectories. This figure of merit will be plotted against the concentration threshold. For the binary case, this simulation will determine the optimum threshold for the detector given a certain chemical power source.

2.3.3. Synthetic Test Case 2: Accuracy in the Estimation of the Background Level and the Expected Position of the Chemical Source

Here, both algorithms (binary and analog) are compared and share the same 10 random trajectories. In this second case, a non-uniform background is used. Results will show the probability evolution at real source location as the exploration time increases. Moreover, the probability maps provided by both approaches after 300 min of exploration time can be compared as well.

To assess the accuracy of a given probability map we compute the expected value of distance to the true source position:

$$E[\mathbf{r} - \mathbf{r}_0] = \sum_{i=1}^{N_c} p(\mathbf{r}_i) \cdot (\mathbf{r}_i - \mathbf{r}_0) \quad (13)$$

where \mathbf{r} is the estimated position of the source, \mathbf{r}_0 is the real source position and $p(\mathbf{r}_i)$ represents the probability of the source being located in cell i .

Based on this metric we calculate the root mean squared error ($RMSE_p$) of the distance from the expected source position to the real source location:

$$RMSE_p = \sqrt{\left(E[\|\mathbf{r} - \mathbf{r}_0\|^2] \right)} = \sqrt{err_x^2 + err_y^2} \quad (14)$$

Alternatively, for a richer description we can decompose the total error in both coordinates: err_x and err_y . Alternative, though less informative, figures of merit are the Euclidean distance (D) between the maximum of the likelihood map and the real source position, and the probability at the real position.

The mean background map recovered for the concentration-based algorithm is quantitatively compared to the designed background distribution by using the root mean squared error ($RMSE_b$) as a figure of merit:

$$RMSE_b = \sqrt{\frac{\sum_{i=1}^{N_c} (B_i - \hat{B}_i)^2}{N_c}} \quad (15)$$

where B_i is the designed mean background level at cell i ; and \hat{B}_i is the estimated mean background level. Results are obtained for two background distributions (maximum mean values: 0.05 ppm and 0.45 ppm). Both background concentration maps were the same except for a scale factor. The binary-based approach was tested using the optimum threshold level determined from case 1 and the concentration-based approach, using the exact source strength.

2.3.4. Synthetic Case 3: Influence of the Source Strength on the Concentration Based Algorithm

Finally, the third simulation shows the influence of the source strength in the overall performance of the concentration-based approach. Since it is difficult to know the source strength in advance, the performance of the algorithm has been assessed assuming different source strengths across more than two orders of magnitude in the range between 0.1 g/s and 30 g/s (the real value in the scenario simulator being 2.90 g/s). Results are shown for the same background distributions as in case 2 and for 10 random trajectories.

2.4. Scenario, Chemical Source Emission and Autonomous Vehicle Description for Real Experiments

For real experiments, all measurements were performed within an exploration arena built with polystyrene panels. Dimensions of the tunnel were 5 m \times 3.5 m \times 1.8 m (length \times width \times height). One fan was introducing air into the room and, on the opposite wall, three fans extracted air from the room. All of them were installed at a height of 0.9 m above the floor and generated a highly a turbulent airflow. Fans used were helical-wall type with a diameter of 30 cm and a maximum speed of 1300 rpm.

The supply of volatile acetone was carried out employing two 10-mL syringes. These syringes were filled with liquid acetone and assembled on a syringe pump KDS-200 (KD Scientific, Holliston, MA, USA) which was programmed to deliver a controlled liquid flow of 150 μ L/min during 60 min. Thus, the liquid acetone was falling at a constant rate over a plate heated to a temperature sufficient for the acetone to be evaporated immediately. The source emission of volatile acetone was placed at one side of the exploration arena just in front of the fan that introduced air in the tunnel. Algorithms for the generation of the probability maps operated over a grid of $N_c = 70$ cells with dimensions 0.5 m \times 0.5 m each one.

Input data for the algorithms were registered by an autonomous vehicle which was constructed on a metal structure where different components were assembled, such as: two DC motors that allow mobility of the vehicle, an UTM-30LX USB laser range finder sensor (HOKUYO, Osaka, Japan) for vehicle auto-location and navigation; and a Windsonic RS232 anemometer (Gill Instruments, Lymington, England) to register the wind speed and direction. The autonomous vehicle was also equipped with a photo ionization detector (PID, ppbRAE 3000, RAE Systems, San Jose, CA, USA) for measurement of the concentration of volatile compounds. All these components were controlled by an onboard computer. The vehicle could operate autonomously sampling its relative position, the wind speed and direction, and the volatile compound concentration at one measurement every second approximately. The vehicle was programmed to move in a straight line until a wall is found, then the vehicle rotates a random angle, and a new straight trajectory is described until a new wall is found. The vehicle speed was set to 0.2 m/s. Images of the robot, the exploration room and the chemical source can be found in the supplementary materials.

3. Results and Discussion

In this section, we will introduce the results for the three cases already described, as well as the results for the real experiments. Finally, we will discuss the computational cost of the presented algorithm. Data are presented as 50th [5th–95th] percentiles unless otherwise specified.

3.1. Results of Algorithms Evaluation for Synthetic Experiments

3.1.1. Synthetic Case 1

The overall performance of the binary-based approach as the concentration threshold is changed is studied in synthetic case 1. The results for two different background levels: the former with a mean value of 0.05 ppm (Figure 3a) and the latter with a mean value of 0.45 ppm (Figure 3b) are shown. The mean probability assigned by the algorithm at the real source location is depicted as a function of the concentration threshold. It can be seen in the figure that there is a different optimal concentration threshold depending on the background level. As expected, the optimum value is shifted to higher thresholds as the background level is increased. These optimal values have been found to be 0.15 ppm and 1.48 ppm, respectively. Relative to the background level, setting the threshold too low produces a high ratio of false alarms leading the algorithm to failure; on the other hand, setting the threshold too high could lead to an increase in false negatives with abnormal concentrations considered as non-detection events and a consequent worsening of the overall performance. This can be critical if the source to be detected is weak –this being equivalent to the case with a high background level. In Figure 3b, it is observed that setting the threshold either too low or too high causes the algorithm to fail, since the probability assigned to the real source location is below the equiprobable value of $1/N_c$ assigned initially to every cell.

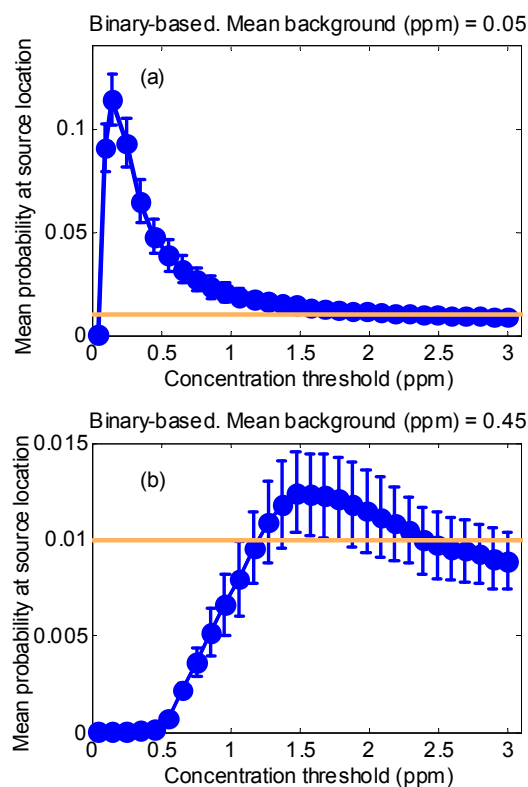


Figure 3. Dependency of the assigned mean probability (averaged over 10 different random trajectories) with the concentration threshold at source location by the binary-based approach. (a) Low background level (mean = 0.05 ppm and SD = 0.03 ppm). (b) High background level (mean = 0.45 ppm and SD = 0.27 ppm). The orange line shows the initial equiprobable value ($1/N_c$) assigned to every cell.

The main problem using the binary-based approach is that the threshold needs to be set arbitrarily if no information about the background is available and this background can be different in various areas within the exploration zone. A priori, we do not know whether the threshold is too high or too low, but even if we knew this, the background could evolve over space and time and the threshold

would need to be adjusted continuously. The concentration-based algorithm removes the necessity of any threshold because, instead of adjusting the threshold level, our approach builds a background model for each cell. This background model allows us to distinguish between the background and the plume without using any threshold and is updated recursively.

3.1.2. Synthetic Case 2

We will first analyze the results for the case of a weak interfering background having a maximum mean value of 0.05 ppm. Figure 4 shows the performance of the binary-based and concentration-based algorithms in source localization by considering the evolution of the figures of merit. Mean probability (averaged over all trajectories) increases throughout the exploration time (Figure 4a) in both approaches.

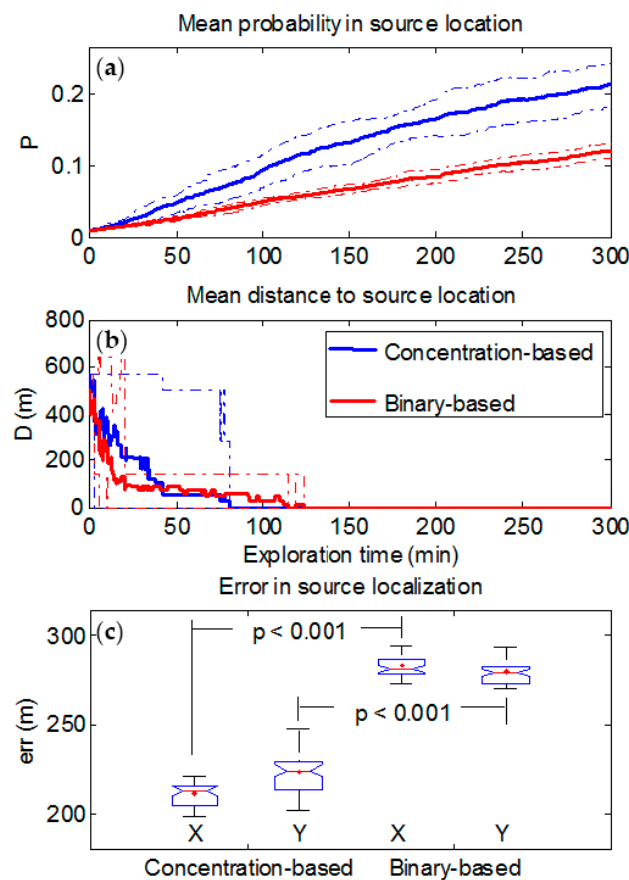


Figure 4. Performance of binary-based and concentration-based algorithms in source localization with a maximum mean background level of 0.05 ppm. (a) Mean probability averaged over the ten random trajectories (thick solid line) and confidence intervals within two standard deviations (thin dashed lines). (b) Mean Euclidean distance D between the cell with the highest probability value and the real source location (thick solid line) and both the maximum and the minimum values (thin dashed lines). (c) error in source localization in both X and Y directions at the end of exploration time; p -values between both approaches by using the Wilcoxon rank test have been indicated.

The Wilcoxon nonparametric rank test (not shown) was applied at each time step to test the null hypothesis of no difference between the median of both populations (i.e., those formed by the 10 probabilities values in each of the two algorithms). Approximately 15 min after the starting time, the increase in probabilities for the concentration-based approach become statistically significant ($p < 0.05$) in comparison with the probabilities for the binary-based approach which did not change until the end of exploration time. Figure 4b shows the mean Euclidean distance D between the cell

with the highest probability value and the real source location. For both approaches, mean distance decreases as the exploration time increases. In average, the binary-based approach converges to the real source location slower (125 min. approximately) than the concentration-based one (80 min. approximately), and with lower probability value as it can be observed in Figure 4-top. Moreover, the errors in source localization algorithms in both X and Y directions of the grid (Figure 4-bottom) show that the concentration-based algorithm performs significantly better ($p < 0.001$), $err = 213$ [199 – 222] m and $err = 225$ [202 – 248] m for X and Y directions respectively, in comparison with the binary approach (281 [274 – 294] and 279 [271 – 294]), even though the binary-based algorithm has been tested with the optimum concentration threshold (0.15 ppm).

Figure 5 shows the comparison between the mean probability maps (averaged over all trajectories) provided by both algorithms after 300 min of random exploration. The probability assigned to the real source location is higher using the concentration-based approach ($P = 0.215$ [0.156 – 0.261] in Figure 5b) as compared to the binary-based approach ($P = 0.124$ [0.102 – 0.131] in Figure 5a). Moreover, the probability is spread among a lower number of cells in the wind direction.

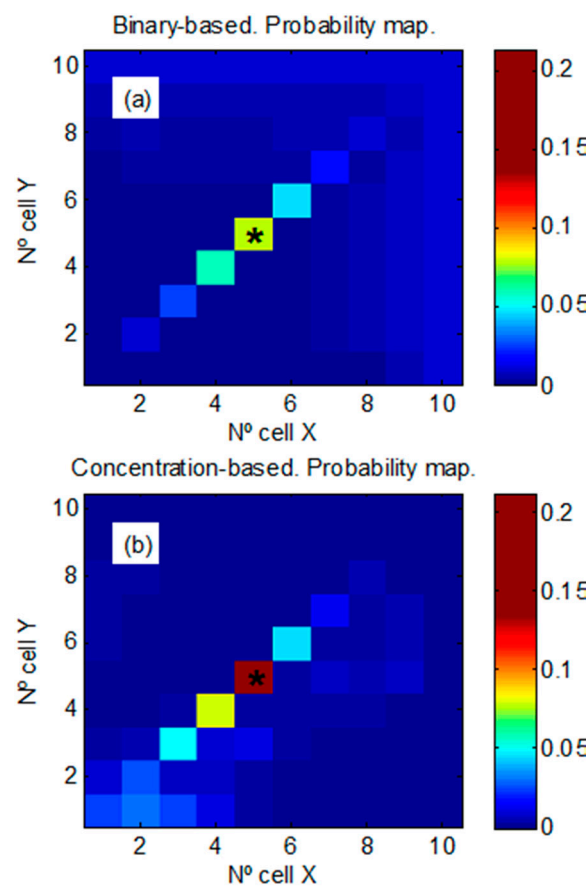


Figure 5. Mean probability maps (averaged over all trajectories) after 300 min of random exploration with a maximum mean background level of 0.05 ppm. Source location at (5, 5) is indicated with an asterisk (*). (a) Binary-based approach. (b) Concentration-based approach.

The estimated background distribution (not shown) by the concentration-based approach is similar to the designed one after 300 min of random exploration, $RMSE_b = 0.010$ [0.009 – 0.012] m. Moreover, it was observed that after 50 min of exploration time the $RMSE_b$ values remained approximately constant.

With a high background level (0.45 ppm), which is the same as saying that the source strength is small compared to the background, Figure 6 shows similar results to those shown in Figure 4. However,

now the differences between the performances of both approaches are higher. Mean probability at real source location (Figure 6a) for the concentration-based approach increases as exploration time increases, but more slowly compared to Figure 4a. In contrast to that observed in Figure 4a for the binary-based approach, the mean probability is lower and remains approximately constant throughout the exploration time. Here, the Wilcoxon test is also used to assess the statistical difference between both approaches. Differences observed become statistically significant ($p < 0.05$) after 15 min of exploration approximately. Moreover, it is observed in Figure 6b that the mean distance for the binary-based algorithm does not converge to the real source location. The errors accounted (Figure 6c) in both the X ($err = 246$ [239 – 265] m) and the Y ($err = 260$ [254 – 277] m) directions are higher than the case of a low background and confirm that the performance of the concentration-based approach is significantly more robust than the binary-based (337 [331 – 338] m and 331 [326 – 341] m for X and Y directions respectively) in the presence of a high background level, although the latter has been tested with its optimum concentration threshold.

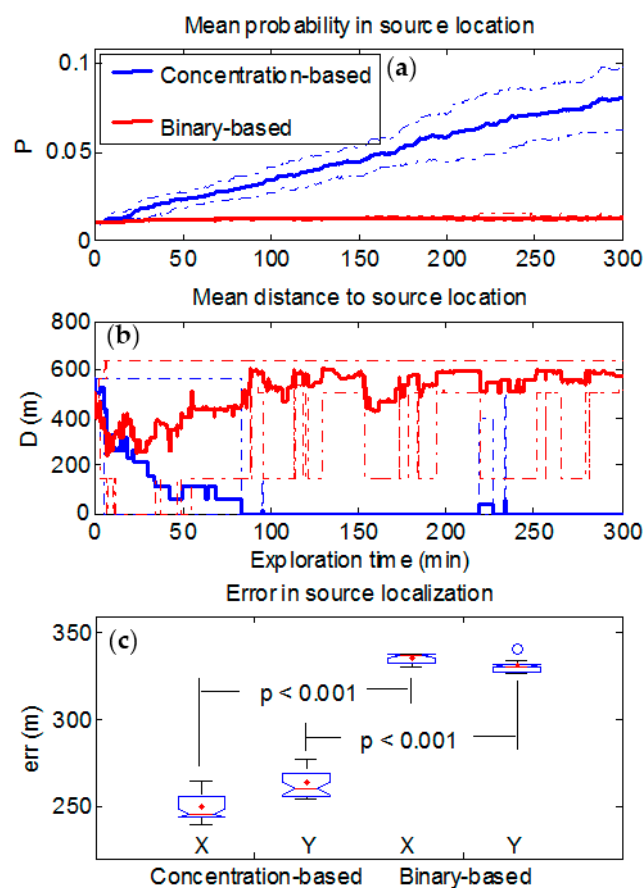


Figure 6. Performance of binary-based and concentration-based algorithms in source localization with a maximum mean background level of 0.45 ppm. (a) Mean probability averaged over the ten random trajectories (thick solid line) and confidence intervals within two standard deviations (thin dashed lines). (b) Mean Euclidean distance D between the cell with the highest probability value and the real source location (thick solid line) and both the maximum and the minimum values (thin dashed lines). (c) error in source localization in both X and Y directions at the end of exploration time; p -values between both approaches by using the Wilcoxon rank test have been indicated.

The binary-based approach performs well under the assumption that no false alarms arise. This is shown in Figure 4 where, due to a low background level, the number of missed detections and false alarms arising from the background are small, and the binary-based algorithm performs slightly

worse compared to the concentration-based approach if the optimum threshold can be identified. Nevertheless, such an assumption is far from the truth in a real scenario where pollution and interfering substances are expected to be found. Additionally, this can also be the case when the source to be detected is weak. This is shown in Figure 6 where, due to a high background level (or weak source), the number of false alarms is higher, thus forcing setting the threshold higher, which leads in turn to an increase in the number of missed detections and worsens dramatically the performance of the binary-based approach.

Figure 7 shows the probability maps obtained with a high background level. It is seen that the threshold is very high, which minimizes the number of false alarms but increases the number of missed detections, thus the binary-based estimation is very uncertain at the source location (Figure 7a). Therefore, the probability is spread over the cells in the grid, decreasing the probability at real source location ($P = 0.013$ [0.010 – 0.014]). Figure 7b shows the robustness of our algorithm which minimizes false alarms and missed detections. The concentration-based algorithm tends to increase the probability at the real source location while the vehicles are performing random exploration. After 300 min of random exploration, the probability assigned to real source location was $P = 0.083$ [0.047 – 0.101]. In this approach, false alarms arising from the background can correctly be assigned lower weights in the probability calculations because the algorithm has created a background model per each cell and a dispersion model for the plume. Additionally, these models allow minimizing the number of missed detections.

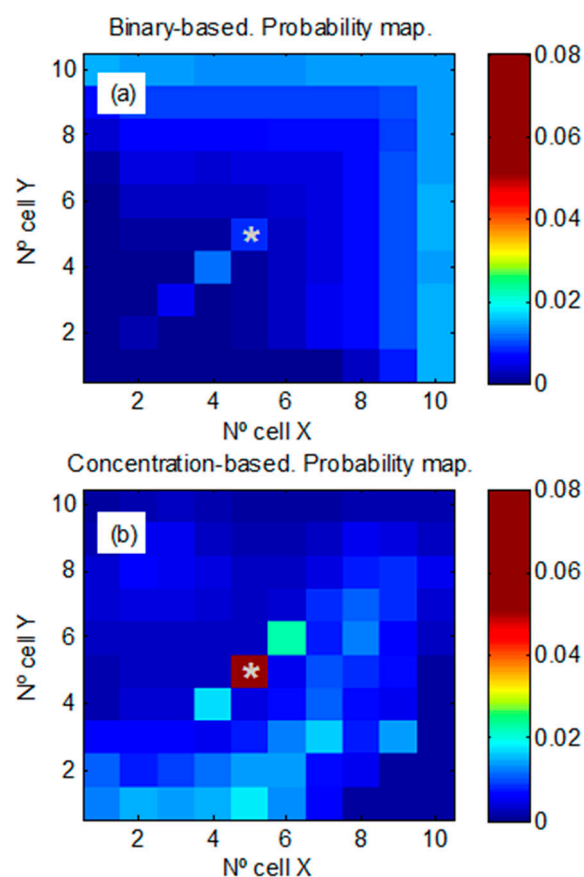


Figure 7. Mean probability maps (averaged over all trajectories) after 300 min of random exploration with a maximum mean background level of 0.45 ppm. Source location at (5, 5) is indicated with an asterisk. (a) Binary-based approach. (b) Concentration-based approach.

Concerning the estimation of the background distribution, the concentration-based approach was also able to recover it properly, $RMSE_b = 0.089$ [0.080 – 0.100] m. Like in the case of a low background, the $RMSE_b$ remained approximately constant after 50 min, albeit, it was slightly higher.

3.1.3. Synthetic Case 3

The results concerning the sensitivity to the correct estimation of the chemical source strength are displayed in Figure 8. Results show a decrease in the overall performance as the assumed source strength deviates from the real value (2.90 g/s). It is observed (Figure 8a) that the selection of the source strength becomes more critical when the background level is low (or the source is potent compared to this background), but, if the source strength could be estimated with errors (to within at most a factor of 4), the concentration-based approach performs much better. When the source to be detected is weak (or the background is comparatively high, Figure 8b), the selection of the source strength is not so critical in the range studied. It can be observed that the concentration-based approach is more robust against false alarms and missed detections (even for source strength estimated with errors larger than two orders of magnitude), as compared to the binary-based approach which performs badly even setting its optimum threshold.

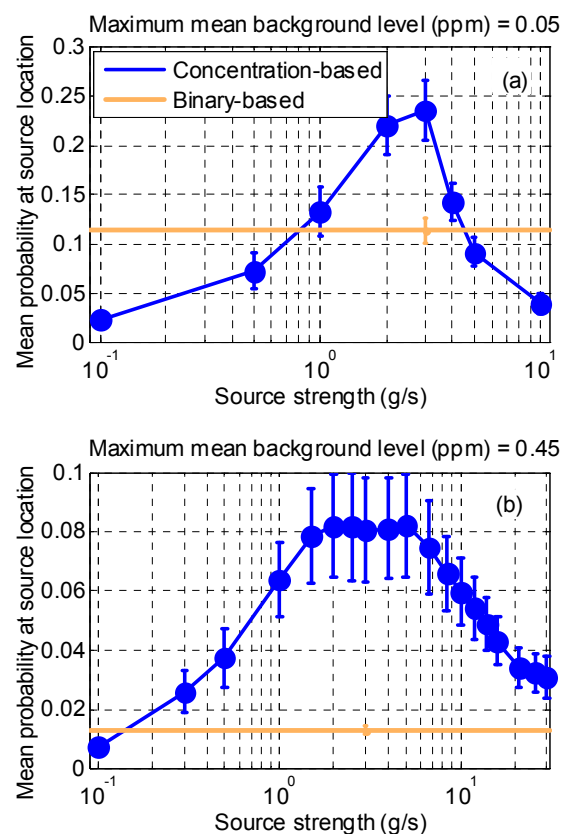


Figure 8. Mean probability (averaged over 10 trajectories) at real source location after 300 min of random exploration as a function of the source strength assumed by the concentration-based approach. Error bars show confidence levels within two standard deviations. (a) Results with a maximum mean background level of 0.05 ppm. (b) Results with a maximum mean background level of 0.45 ppm.

It is important to say that, although the binary-based algorithm works without assuming explicitly any source strength, setting an optimum threshold is only possible when collecting real measurements which implicitly contain information about the source strength and the background. However, this optimum threshold might be different depending on the explored cell. In the case of the concentration-based approach, the background is estimated by the algorithm and, if this background

is low, the source strength needs to be known to within one order of magnitude. If the background is high, the algorithm behaves more robustly in the range studied as compared to the binary case except in the first case (the assumed source strength is very small).

3.2. Results for the Real Experiments

First we will give some details concerning the wind distribution within the exploration arena. Figure 9 (top) shows a characterization of the behavior of the wind within the designed scenario for one of the experiments that was carried out. The median wind speed was 0.28 [0.08 – 1.12] m/s. Also, there was a predominant wind direction although with a scatter on the direction angle of approximately 60 degrees. Similar behavior of the wind was observed in all the experiments. The map of concentrations showed in Figure 9 (bottom) is consequence in part of such airflow generated.

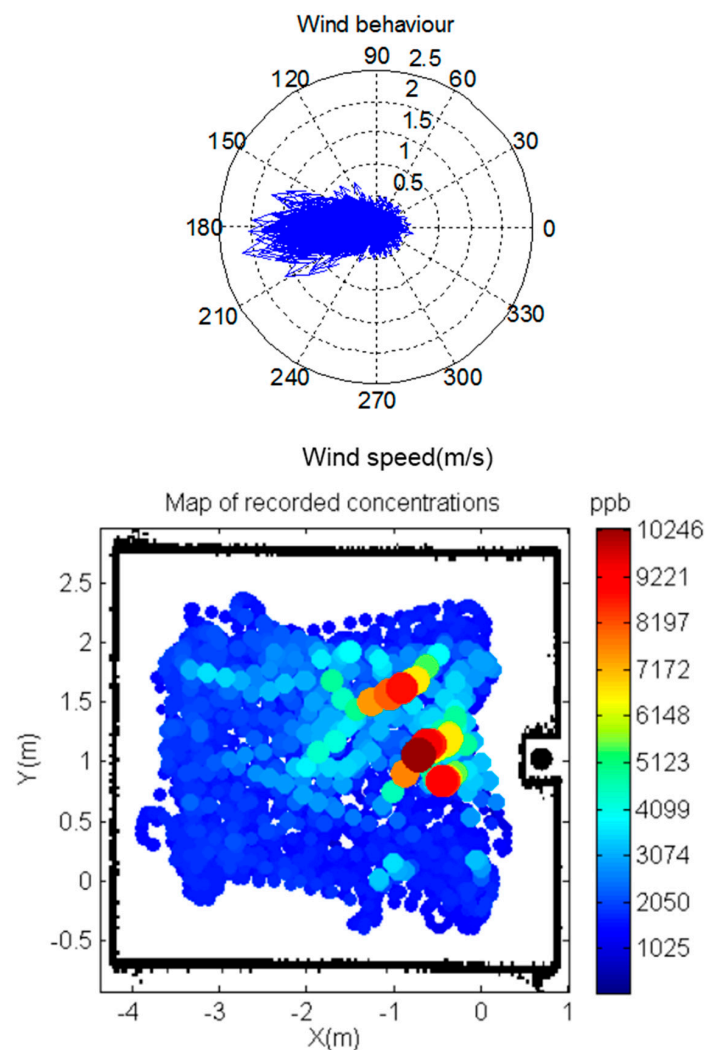


Figure 9. (Top) a representative wind distribution recorded within the designed scenario during a random exploration maneuver. Both, the direction (degrees) and the speed (m/s) of the wind are indicated. (Bottom) map of concentrations recorded with the PID. Source location at (10, 4) is indicated by a black circle.

The averaged probability maps (over the 10 experiments) are shown in Figure 10. The mean probability assigned by each algorithm to the real source location is practically identical, $P = 0.045$ [0.030 – 0.076] and $P = 0.050$ [0.028 – 0.084] for the concentration-based and for the binary-based respectively. However, the probability map for the binary-based algorithm (Figure 10a)

exhibits more variance in the localization of the real source position, as it is observed in Figure 10a. For the concentration-based algorithm (Figure 10b), relatively stable higher values of probability around the real source location are obtained. The Wilcoxon rank test is also used to compare the errors obtained for each algorithm. In the X direction, the error for the concentration-based algorithm (2.14 [2.09 – 2.35]) m. is lower than for the binary-based one (2.30 [2.03 – 2.55]) m., but without statistical significance. In the Y direction, the error for the concentration-based (0.91 [0.85 – 0.95]) m. is also lower than for the binary-based (0.97 [0.86 – 1.03]) m. and in this case is statistically significant (see Figure 11).

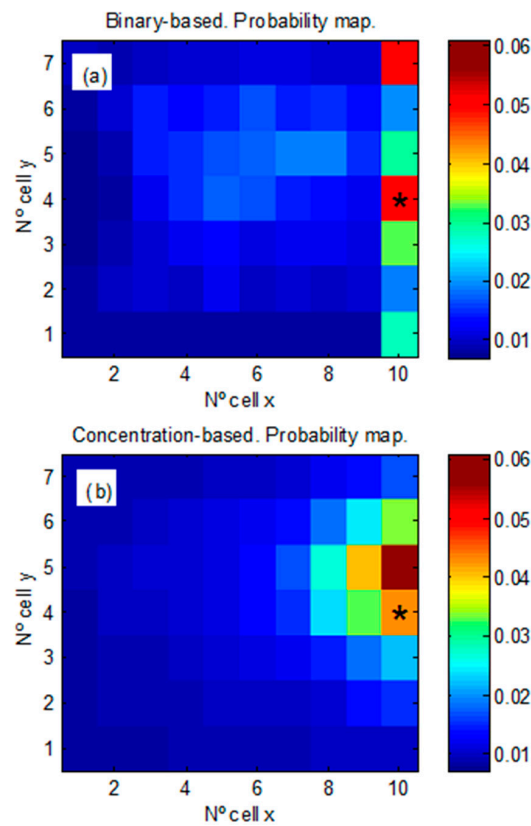


Figure 10. Mean probability maps at the end of the exploration time for the set of real experiments. Source location at (10, 4) is marked with an asterisk. (a) Binary-based approach. (b) Concentration-based approach.

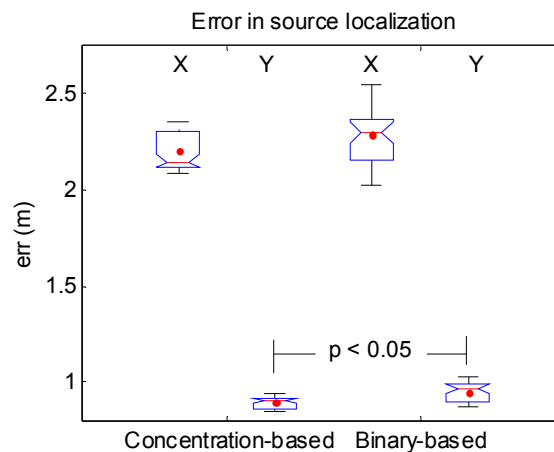


Figure 11. Errors in source localization in both the X and the Y directions for both algorithms. Statistically significant difference is indicated.

Results obtained here are in the same line as those obtained for the simulation experiments but the results are not as clear. In the designed scenario, it is likely that some of the assumptions of our algorithm (see Table A2) are not fully satisfied. The wind was not completely uniform in the room: it was much stronger near the source than at the opposite wall and there were some recirculation at the lateral walls. As a consequence the real distribution of the time-averaged concentration deviated from Gaussianity. We think it might be a reason to explain the slight differences observed when comparing both algorithms, in contrast to the strong differences observed in the performance of both algorithms for the simulation experiments.

3.3. Computational Cost

Finally, we have studied the computational cost of the algorithm. It is true that the computational cost is higher than the binary approach, mainly due to need of computing the convolution between the PDF of the background and the PDF of the plume for every cell. The cost is about three times the cost of the binary approach in our current implementation (without optimization). However, our studies have confirmed that the computational cost of each iteration in the algorithm increases very slowly with the number of cells. Moreover, the time to locate the source increases in a linear way approximately with the number of cells.

4. Conclusions

In the present paper, modifications of a previously described (binary-based) algorithm have been introduced. The original algorithm can be easily extended to work with multiple mobile sensors. All the information from the mobile sensors can be integrated in the algorithm, whatever their positions are. The algorithm only needs to know in which cell the concentration readings were obtained, then a probability map will be recursively updated. Moreover, the mobile sensors do not need to solely perform plume tracking and might be used for other tasks.

Additionally, in a real scenario, pollution and some interfering substances may appear in the background, increasing the number of false alarms. Unlike the binary-based algorithm, which uses a threshold to assess whether a concentration is considered as a detection or non-detection event; our algorithm, based on continuous concentrations, builds a background model to assess whether a concentration comes from the background or from a source located further away. Simulation results show that our algorithm behaves much more robustly in the presence of false alarms and better estimates the real source location.

All concentration readings are considered in our algorithm, incorporating them in a continuous manner instead of just using them as binary detections above a certain threshold. This fact removes the need for a threshold level, thereby reducing the number of false alarms (a background model is estimated) and the number of missed detections thus improving the performance of the algorithm proposed by Pang and Farrell. A sensitivity study regarding robustness of the algorithm against deviations from the true value has been presented. It has been shown, the results improve significantly using the concentration-based algorithm if the source strength can be estimated.

Finally, experiments with real data have shown that the concentration-based algorithm seems to perform slightly better than the binary-based one, confirming our results obtained for the simulation experiments. A critical point in our proposal is that the algorithm assumes that the source strength is known. Thus, estimating the source strength would be a promising direction for future research. Results from real experiments show that when the plume dispersion model hypothesis do not hold, both algorithm are still able to perform the estimation tasks although the differences diminish.

The described algorithm is available in MATLAB code from the authors under request.

Supplementary Materials: The supplementary materials are available online at <http://www.mdpi.com/1424-8220/17/4/904/s1>. Images of the robot, the exploration arena with the fans and the used chemical source are available as supplementary materials.

Acknowledgments: The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 217925: LOTUS: Localization of Threat Substances in an Urban Society. This work was partially funded by the Spanish MINECO program, under grants TEC2011-26143 (SMART-IMS), TEC2014-59229-R (SIGVOL), and BES-2015-071698 (SEVERO-OCHOA). The Signal and Information Processing for Sensor Systems group is a consolidated Grup de Recerca de la Generalitat de Catalunya and has support from the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (expedient 2014-SGR-1445). This work has received support from the Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya and the European Social Fund (ESF). Additional financial support has been provided by the Institut de Bioenginyeria de Catalunya (IBEC). IBEC is a member of the CERCA Programme/Generalitat de Catalunya.

Author Contributions: S.M. designed the study and provided the main ideas behind the concentration based algorithm. V.P. programmed the simulator, the implementation of both algorithms and wrote a first version of the manuscript. V.P., R.M. and J.B. analyzed the signals and ran the algorithm for several test cases and for the real case signals. D.M., M.T. and J.P. designed and programmed the autonomous robot. R.M., J.M.J.-S., D.M. and M.T. mounted the exploration arena and carried out the real experiments. R.M. and S.M. contributed to the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A. Supplementary Tables

Table A1. Summary of the notation used for the concentration-based approach.

N_c	Number of Cells in the Grid Area
L_x	Length of each cell along the x -axis
L_y	Length of each cell along the y -axis
c	Instantaneously measured concentration
c_b	Concentration contribution due to the background
c_p	Concentration contribution due to the chemical plume
\bar{c}	Mean concentration at a fixed location
q	Source strength or release rate
U_a	Mean wind speed in the downwind direction
σ_y	Diffusion coefficient in the crosswind direction
σ_z	Diffusion coefficient in the vertical direction
\hat{c}	Array of all possible instantaneous concentrations c
γ	Intermittency factor related to the chemical plume
M	Mean concentration of a series of readings at a fixed location
σ_{SD}	Standard deviation of a series of readings at a fixed location
N_b	Number of readings stored in the concentration buffer per each sensor
c_j	Measured concentration within cell j
A_i	Event "there is a chemical source within cell i "
$B(t_k)$	Sequence of measured concentrations along the trajectory of the robots until time t_k
$P(A_i)$	Prior probability of the presence of a chemical source within cell i
$P(c_j A_i)$	Probability that the measurement within cell j is due to the addition of the background at cell j and a chemical plume due to a source within cell i
$P(c_j \bar{A}_i)$	Probability that the measurement of chemical at cell j is not due to a source emitting at cell i , thus c_j is due to the current background at cell j
$S_{ij}(t_k) \equiv P(A_i c_j)$	Probability of having a source in cell i given that a certain amount of chemical was measured at cell j at time t_k
$\beta_{ij}(t_k)$	Normalized probability (over all cells) of having a chemical source within cell i based on a single measured concentration within cell j at time t_k
$\alpha'_{ij}(t_k)$	Normalized probability (over all cells) of having a chemical source within cell i based on the sequence of measured concentrations along the trajectory of the robots (index j) until time t_k

Table A2. Main assumptions of the concentration-based algorithm.

For the Plume
Gaussian distribution for the time-averaged concentration concentration fluctuations governed by turbulences continuous release, source strength known $q = 2.90$ g/s one uniquely source at the position (400 m, 400 m) height, $z = 2$ m
For the Background
background is always present background relatively constant for the exploration time mean background level can change from one cell to another no intermittency
Other Assumptions
uniform wind field over the exploration time, speed $U_a = 2.5$ m/s, and 45° direction neutral atmospheric stability no deposition of the substance on surfaces cells are equally likely to contain the source at time t_0 height of sensors is 2 m response-time of sensors faster than the typical 10min time-average of GPM

References

1. Kowadlo, G.; Russell, R.A. Robot Odor Localization: A Taxonomy and Survey. *Int. J. Robot. Res.* **2008**, *27*, 869–894. [[CrossRef](#)]
2. Webster, D.R.; Volyanskyy, K.Y.; Weissburg, M.J. Bioinspired algorithm for autonomous sensor-driven guidance in turbulent chemical plumes. *Bioinspir. Biomimetics* **2012**, *7*, 3. [[CrossRef](#)] [[PubMed](#)]
3. Meng, Q.-H.; Yang, W.-X.; Wang, Y.; Li, F.; Zeng, M. Adapting an ant colony metaphor for multi-robot chemical plume tracing. *Sensors* **2012**, *12*, 4737–4763. [[CrossRef](#)] [[PubMed](#)]
4. Russell, R.A.; Bab-Hadiashar, A.; Shepherd, R.L.; Wallace, G.G. A comparison of reactive robot chemotaxis algorithms. *Robot. Autom. Syst.* **2003**, *45*, 83–97. [[CrossRef](#)]
5. Webster, D.R. Structure of turbulent chemical plumes. In *Trace Chemical Sensing of Explosives*; Woodfin, R.L., Ed.; John Wiley & Sons: Hoboken, NJ, USA, 2007; pp. 109–129.
6. Jatmiko, W.; Mursanto, P.; Kusumoputro, B.; Sekiyama, K.; Fukuda, T. Modified PSO algorithm based on flow of wind for odor source localization problems in dynamic environments. *J. WSEAS Trans. Syst.* **2008**, *7*, 106–113.
7. Ishida, H.; Suetsugu, K.; Nakamoto, T.; Moriizumi, T. Study of autonomous mobile sensing system for localization of odor source using gas sensors and anemometric sensors. *Sens. Actuators A Phys.* **1994**, *45*, 153–157. [[CrossRef](#)]
8. Vergassola, M.; Villermaux, E.; Shraiman, B.I. Infotaxis as a strategy for searching without gradients. *Nature* **2007**, *445*, 406–409. [[CrossRef](#)] [[PubMed](#)]
9. Lilienthal, A.; Duckett, T. Building gas concentration gridmaps with a mobile robot. *Robot. Autom. Syst.* **2004**, *48*, 3–16. [[CrossRef](#)]
10. Stachniss, C.; Plagemann, C.; Lilienthal, A.J. Learning gas distribution models using sparse Gaussian process mixtures. *Auton. Robot.* **2009**, *26*, 187–202. [[CrossRef](#)]
11. Lilienthal, A.J.; Reggente, M.; Trincavelli, M.J.; Blanco, L.; Gonzalez, J. A Statistical Approach to Gas Distribution Modelling with Mobile Robots—The Kernel DM + V Algorithm. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, USA, 10–15 October 2009; pp. 570–576.
12. Farrell, J.A. Plume Mapping via hidden markov methods. *IEEE Trans. Syst. Man Cybern. Part B* **2003**, *33*, 850–863. [[CrossRef](#)] [[PubMed](#)]
13. Farrell, J.A.; Pang, S.; Li, W. Chemical Plume Tracing via an Autonomous Underwater Vehicle. *IEEE J. Ocean. Eng.* **2005**, *30*, 428–442. [[CrossRef](#)]
14. Pang, S.; Farrell, J.A. Chemical Plume Source Localization. *IEEE Trans. Syst. Man Cybern. Part B* **2006**, *36*, 1068–1080. [[CrossRef](#)]

15. Kang, X.; Li, W. Moth-inspired plume tracing via multiple autonomous vehicles under formation control. *Adapt. Behav.* **2012**, *20*, 131–142. [[CrossRef](#)]
16. La, H.M.; Sheng, W.; Chen, J. Cooperative and Active Sensing in Mobile Sensor Networks for Scalar Field Mapping. *IEEE Trans. Syst. Man Cybern.* **2015**, *45*, 1–12. [[CrossRef](#)]
17. La, H.M.; Shen, W. Distributed Sensor Fusion and Scalar Field Mapping Using Mobile Sensor Networks. *IEEE Trans. Syst. Man Cybern.* **2013**, *45*, 766–778.
18. Choi, J.; Oh, S.; Horowitz, R. Distributed learning and cooperative control for multi-agent systems. *Automatica* **2009**, *45*, 2802–2814. [[CrossRef](#)]
19. Pomareda, V.; Marco, S. Chemical Plume Source Localization with Multiple Mobile Sensors using Bayesian Inference under Background Signals. In Proceedings of the International Symposium on Olfaction and Electronic Nose, New York, NY, USA, 2–5 May 2011; pp. 149–150.
20. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1984.
21. Turner, B. *Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modeling*; CRC Press: Boca Raton, FL, USA, 1994.
22. Beychok, M.R. *Fundamentals of Stack Gas Dispersion*, 4th ed.; Beychok, M.R.: New York, NY, USA, 2005.
23. Sutton, O.G. The problem of diffusion in the lower atmosphere. *Q. J. R. Meteorol. Soc.* **1947**, *73*, 257–281. [[CrossRef](#)]
24. Crimaldi, J.P.; Wiley, M.B.; Koseff, J.R. The relationship between mean and instantaneous structure in turbulent passive scalar plumes. *J. Turbul.* **2002**, *3*, 1–24. [[CrossRef](#)]
25. Webster, D.R.; Weissburg, M.J. Chemosensory guidance cues in a turbulent chemical odor plume. *Limnol. Oceanogr.* **2001**, *46*, 1034–1047. [[CrossRef](#)]
26. Bakkum, E.A.; Duijm, N.J. Chapter 4: Vapour cloud dispersion. In *Methods for the Calculation of Physical Effects 'Yellow Book'*, 3rd ed.; van den Bosch, C.J.H., Weterings, R.A.P.M., Eds.; Ministerie van verkeer en waterstaat: The Hague, The Netherlands, 2005; pp. 1–139.
27. Farrell, J.A.; Murlis, J.; Long, X.; Li, W.; Cardé, R.T. Filament-based atmospheric dispersion model to achieve short time-scale structure of odor plumes. *Environ. Fluid Mech.* **2002**, *2*, 143–169. [[CrossRef](#)]
28. Yee, E.; Wang, B.-C.; Lien, F.-S. Probabilistic model for concentration fluctuations in compact-source plumes in an urban environment. *Bound.-Layer Meteorol.* **2009**, *130*, 169–208. [[CrossRef](#)]
29. Yee, E.; Biltoft, C.A. Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles. *Bound.-Layer Meteorol.* **2004**, *111*, 363–415. [[CrossRef](#)]
30. Yee, E.; Chan, R. A simple model for the probability density function of concentration fluctuations in atmospheric plumes. *Atmos. Environ.* **1997**, *31*, 991–1002. [[CrossRef](#)]
31. Jones, A.R.; Thomson, D.J. Simulation of time series of concentration fluctuations in atmospheric dispersion using a correlation-distortion technique. *Bound.-Layer Meteorol.* **2006**, *118*, 25–54. [[CrossRef](#)]
32. Hahn, I.; Brixey, L.A.; Wiener, R.W.; Henkle, S.W.; Baldauf, R. Characterization of traffic-related PM concentration distribution and fluctuation patterns in near-highway urban residential street canyons. *J. Environ. Monit.* **2009**, *11*, 2136–2145. [[CrossRef](#)] [[PubMed](#)]
33. Soriano, C.; Baldasano, J.M.; Buttler, W.T.; Moore, K.R. Circulatory patterns of air pollutants within the Barcelona air basin in a summertime situation: lidar and numerical approaches. *Bound.-Layer Meteorol.* **2000**, *98*, 33–55. [[CrossRef](#)]
34. Martín, M.; Cremades, L.V.; Santabàrbara, J.M. Analysis and modelling of time series of surface wind speed and direction. *Int. J. Climatol.* **1999**, *19*, 197–209. [[CrossRef](#)]
35. Gijbels, I.; Hubert, M. Robust and nonparametric statistical methods. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, 1st ed.; Brown, S.D., Tauler, R., Walczak, B., Eds.; Elsevier: Amsterdam, The Netherlands, 2009; pp. 189–211.
36. Papoulis, A. Chapter 8: Sequences of random variables. In *Probability, Random Variables, and Stochastic Processes*; McGraw-Hill: New York, NY, USA, 1984.

